

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:

Директор



А. В. Замятин

20 23 г.

Рабочая программа дисциплины

Технологии высокопроизводительной обработки больших данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки :

Интеллектуальный анализ больших данных

Форма обучения

Очная

Квалификация

Магистр

Год приема

2023

Код дисциплины в учебном плане: Б1.В.ДВ.02.01.01

СОГЛАСОВАНО:

Руководитель ОП

А.В. Замятин

Председатель УМК

С.П. Сущенко

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

- УК-1 – способность осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий;
- ПК-5 – способность управлять получением, хранением, передачей, обработкой больших данных.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИУК-1.3 Предлагает и обосновывает стратегию действий с учетом ограничений, рисков и возможных последствий.

ИУК-1.2 Осуществляет поиск, отбор и систематизацию информации для определения альтернативных вариантов стратегических решений в проблемной ситуации.

ИУК-1.1 Выявляет проблемную ситуацию, на основе системного подхода осуществляет её многофакторный анализ и диагностику.

ИПК-5.1 Осуществляет мониторинг и оценку производительности обработки больших данных.

ИПК-5.2 Использует методы и инструменты получения, хранения, передачи, обработки больших данных.

ИПК-5.3 Разрабатывает предложения по повышению производительности обработки больших данных.

2. Задачи освоения дисциплины

– Обучить студентов различным методам организации высокопроизводительных вычислений, сформировать навыки использования различных программных инструментов для организации параллельных вычислений.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор. Дисциплина входит в Прикладной модуль.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Первый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 16 ч.

-лабораторные: 32 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Раздел 1. Введение

- 1.1. Актуальность, базовая терминология и тенденции развития.
- 1.2. Параллелизм компьютерных вычислений.

Раздел 2. Архитектура вычислительных систем

- 2.1. Архитектура высокопроизводительных вычислительных систем
- 2.2. Классификация вычислительных систем

Раздел 3. Облачные технологии, их свойства и типы

Раздел 4. Технология вычислений MapReduce

Раздел 5. Распределённые файловые системы

Раздел 6. Программирование для высокопроизводительных вычислений.

- 6.1. Методы программирования для высокопроизводительных вычислений
- 6.2. Методология проектирования параллельных алгоритмов

9. Текущий контроль по дисциплине

Текущий контроль освоения дисциплины осуществляется в виде проведения регулярных опросов по пройденному материалу и проверки знаний при допуске к выполнению лабораторных работ. Также текущий контроль успеваемости осуществляется в виде сдачи лабораторных работ.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Промежуточная аттестация по курсу проводится в форме экзамена.

Типовые контрольные вопросы для проведения промежуточной аттестации:

1. Виды параллелизма компьютерных вычислений.
2. Классификацию вычислительных систем Флинна.
3. Классификация MIMD систем.
4. Свойства и типы облачных технологий.
5. Технология распределённых вычислений MapReduce.
6. Реализации MapReduce: сходства и различия.
7. Распределённые файловые системы.
8. Распределённая файловая система Google File System.
9. Распределённая файловая система Hadoop distributed file system.
10. Методология организации параллельных вычислений для SIMD, MIMD архитектур.
11. Показатели качества параллельных методов.
12. OpenMP и MPI: назначение каждой из технологий и их сравнительный анализ.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– S. Srinivasan. Cloud Computing Basics electronic resource. – New York, NY : Springer New York : Imprint: Springer, 2014.

– A Ohri. R for Cloud Computing electronic resource : An Approach for Data Scientists. – New York, NY : Springer New York : Imprint: Springer, 2014.

– Zaigham Mahmood. Cloud Computing: Methods and Practical Approaches. – London : Springer London : Imprint: Springer, 2013.

– К. Ю. Богачев. Основы параллельного программирования. – Москва БИНОМ. Лаборатория знаний, 2015.

– В. П. Гергель. Теория и практика параллельных вычислений : учебное пособие. – Москва : Интернет-Университет Информационных Технологий, 2016.

б) дополнительная литература:

– Christoph Fehling, Frank Leymann, Ralph Retter, Walter Schupeck, Peter Arbitter. Cloud Computing Patterns electronic resource : Fundamentals to Design, Build, and Manage Cloud Applications. – Vienna : Springer Vienna : Imprint: Springer, 2014.

– Xiaolin Li, Judy Qiu. Cloud Computing for Data-Intensive Applications electronic resource. – New York, NY : Springer New York : Imprint: Springer, 2014.

– Zaigham Mahmood. Cloud Computing electronic resource : Challenges, Limitations and R&D Solutions. – New York, NY : Springer New York : Imprint: Springer, 2014.

– А. В. Линева, Д. К. Боголепов, С. И. Бастратов. Технологии параллельного программирования для процессоров новых архитектур. – Нижегородский гос. ун-т им. Н. И. Лобачевского – Москва : Изд-во Московского университет, 2010.

в) ресурсы сети Интернет:

– Облачный сервис <http://www.ncloudtech.ru>

– Статья «Бизнес в облаках. Чем полезны облачные технологии для предпринимателя» <https://kontur.ru/articles/225>

– Батура Т.В., Мурзин Ф.А., Семич Д.Ф. Облачные технологии: основные понятия, задачи и тенденции развития. <http://swsys-web.ru/cloud-computing-basic-concepts-problems.html>

– MapReduce Tutorial https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html

– HDFS Architecture Guide https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

Microsoft Visual Studio Community Edition 2013 (C++), библиотека классов, реализующая технологию MapReduce (например, <https://github.com/cdmh/mapreduce>).

14. Материально-техническое обеспечение

Для материально-технического обеспечения дисциплины требуется наличие компьютерной техники с установленным соответствующим программным обеспечением и другого оборудования, поддерживающего проведение презентаций.

Технические характеристики: Компьютер INTANT i6250_T (i5-6500, 3.2GHz\GA-H110M-S2V, mATX\8ГБ DDR3, 1600МГц\1000ГБ 7200об/мин., 32МБ SATAIII\Microtower 400Вт\DVD±RW\2 x 23,6" АОС I2476VW \кл-ра+мышь). Монитор LCD 23" Philips 234E5QDAB – 13 шт.

Интерактивная доска SMART Board SB480, ультракороткофокусный проектор SMART UF70 с настенным креплением

15. Информация о разработчиках

Дружинин Денис Вячеславович, канд. техн. наук, доцент кафедры теоретических основ информатики ТГУ.